DOCUMENT RESUME

ED 393 892 TM 024 822

AUTHOR DeVaney, Thomas A.; Franks, Melvin E.
TITLE An Examination of the Consistency of the Mathematical Constructs Assessed by Classroom Grades, Portfolios, and Standardized Tests for Elementary Grade Students.
PUB DATE Nov 95
NOTE 17p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Biloxi, MS, November 8-10, 1995).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Black Students; Construct Validity; Educational Assessment; Elementary Education; Elementary School Students; Factor Analysis; Grade 3; Grade 4; Grade 5; Grade 6; *Grades (Scholastic); *Mathematics Tests; *Portfolios (Background Materials); *Standardized Tests; Test Construction; *Test Reliability

ABSTRACT

If teachers are teaching a set of standard content and assessments are made consistently, the relationship between these various assessments and the curriculum should be of interest. The purpose of this study was to explore the consistency of the mathematical measures of three prominent forms of mathematics assessment: standardized tests, portfolios, and classroom grades. The sample used in this study was a nonrandom sample of 50 African-American elementary students in grades 3 through 6. Three factor analyses were conducted to determine a general pattern among the variables for the overall sample and explore the consistency of the overall results for grade 3 students and grade 4 through 6 students. The results give evidence supporting the belief that the standardized test and portfolio measure different constructs and that both forms of assessment exert an influence on classroom assessment. (Contains eight tables and seven references.) (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 393 892

AN EXAMINATION OF THE CONSISTENCY OF THE MATHEMATICAL CONSTRUCTS ASSESSED BY CLASSROOM GRADES, PORTFOLIOS, AND STANDARDIZED TESTS FOR ELEMENTARY GRADE STUDENTS

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
THOMAS A. DEVANEY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

by

Thomas A. DeVaney
Program of Research and Evaluation for Public Schools
Mississippi State University

and

Melvin E. Franks
Program of Research and Evaluation for Public Schools
Mississippi State University

Annual Meeting of the Mid-South Educational Research Association
Biloxi, Mississippi
November, 1995

TM024822

BEST COPY AVAILABLE

Abstract

If teachers are teaching a set of standard content and assessments are made consistently, the relationship between these various assessments and the curriculum should be of interest. The purpose of this study was to explore the consistency of the mathematical measures of three prominent forms of mathematics assessment: standardized tests, portfolios, and classroom grades. The sample used in this study was a nonrandom sample of 50 African-American elementary students in grades 3 through 6. Three factor analyses were conducted to determine a general pattern among the variables for the overall sample and explore the consistency of the overall results for grade 3 students and grade 4 through 6 students. The results give evidence supporting the belief that the standardized test and portfolio measure different constructs and that both forms of assessment exert an influence on classroom assessment.

Introduction

As part of a Magnet Assistance School Program, students in grades three through six were provided mathematics instruction according to the latest NCTM standards with emphasis on solving *real-life* problems. After much discussion, the researchers concluded that teachers were teaching a set of standard concepts as outlined in the school district's mathematics curriculum and that the students were given three different assessments: a national standardized test, a portfolio assessment, and regular classroom quarterly grades. If teachers are teaching a set of standard content and assessments are made consistently, the relationship between these various assessments and the curriculum should be of interest. The purpose of this study was to explore the consistency of the mathematical measures of these three prominent forms of mathematics assessment.

Methodology

Sample

The sample used in this study was a nónrandom sample of 50 African-American elementary students in grades 3 through 6. The sample consisted of 19 males and 31 females. Of the 50 students, 22 were in grade 3, 10 were in grade 4, 14 were in grade 5, and 4 were in grade 6. All of the students in the sample were enrolled in a magnet school program located within a traditional school setting. The program was centered around a multi-cultural theme and was designed to make extensive use of multimedia technology.

Dependent Variables

CTB McGraw Hill Mathematics Portfolio. The CTB Portfolio Assessment is designed to offer school districts access to an easily administered mathematics assessment for grades 3 through 8. The assessment is designed to divide the school year into four equal periods. At the end of each period, students are assigned performance activities that require them to integrate their mathematics knowledge in solving *real-life* problems. Like traditional portfolios, students accumulate their work in portfolio folders that can be kept in the classroom. The CTB Portfolio Assessment includes scoring guidelines that are unique to each activity ("CTB Portfolio", 1992). However, for this study, the scoring was slightly modified and results were reported as the percent correct. Additionally, the assessment was only administered during the last nine week period of the school year.

Stanford Achievement Test. The *Stanford Achievement Test* (SAT), Eighth Edition is a national standardized test that is "designed to reflect what is being taught in schools throughout the country." ("Stanford Achievement", 1990, p. 11) The test was administered during the third nine weeks and the score used in this analysis was the national percentile for each student on the mathematics subtest.

Classroom Grades. Classroom assessment was measured using the teacher reported grades for each of the four nine week periods in the 1993-1994 academic year.

Procedure

The data used in this study came from a larger study involving two magnet school programs. The data for the CTB Mathematics Portfolio was furnished by the school district in the form of computer printouts. The Stanford Achievement Test percentile scores were also

furnished by the district via diskette. The data was confirmed using computer printouts since students received different identification numbers and only fourteen of the students were able to be matched on a computer using their identification numbers. The classroom grades were collected by the researchers with cooperation from the participating school.

Method

In order to determine whether the measures addressed the same constructs, three separate factor analyses were conducted. The first analysis was conducted using the entire sample of 50 students. This analysis was used to determine a general pattern among the assessments. The remaining two analyses explored the patterns for grade 3 students and grade 4 through 6 students separately in order to determine the consistency of the overall factor pattern. For each analysis, a principal components technique was used to extract factors, and the number of factors extracted was determined using the eigenvalue-one criterion. When applicable, the initial factor loading matrices were rotated using the Varimax rotation technique in order to aid in the interpretation of the factors. Using criteria contained in Hair, Anderson, Tatham, and Black (1995) factor loadings larger than +/- 0.50 were considered practically significant.

Results and Discussion

Analyses Involving the Overall Sample

The correlations between each of the mathematics measures are presented in Table 1. Of the fifteen computed correlations, nine were statistically significant at the .05 level. The largest correlation was between the first nine weeks average and the second nine weeks average, $r = .71$,

$p < .05$. The weakest correlation was between the Stanford Achievement Test and the CTB Mathematics Portfolio, $r = .12$, $p > .05$.

Insert Table 1 About Here

The unrotated factor solution for the analysis involving the overall sample is contained in Table 2. The table shows that the strongest associations with factor 1 occurred with the first and second nine weeks average. Additionally, the strongest associations with factor 2 occurred with the third and fourth nine weeks average. Table 2 also provides the final communality estimates for each of the assessments. These values range from a high of .77 for the third nine weeks average to a low of .36 for the Stanford Achievement Test. Finally, the table shows that factor 1 accounts for 41.2% of the variance while factor 2 accounts for an additional 22.5%.

Insert Table 2 About Here

Since two factors were extracted in this analysis, the Varimax rotation technique was applied to the unrotated factor matrix. The results of the rotation technique are contained in Table 3. The results show that four of the six measurements weight significantly on one factor and low on the remaining factor. Although the assessments for the first and second nine weeks weight significantly on factor 1, .65 and .70, respectively, they also provide weight to factor 2, .48 and .51, respectively.

Insert Table 3 About Here

When viewing the rotated factor loadings for the Stanford Achievement Test and the CTB Mathematics Portfolio, it appears that factor 1 represents mathematical knowledge as measured by the Stanford Achievement Test. Similarly factor 2 appears to represent mathematical knowledge defined by the CTB Mathematics Portfolio. Further examination of the quarterly classroom grades shows that the association between the classroom grades and factor 1 increases from the first to third quarter before decreasing in the fourth quarter. The factor loadings for the classroom grades on factor 2 are similar to those for factor 1. There is an increase in association from the first to the second quarter. However, the association between the third quarter and factor 2 decreases substantially before returning to a significant association in the fourth quarter.

Analyses for Students in Grade 3

Table 4 contains the pairwise correlations between each of the mathematics measures. The table shows that thirteen of the fifteen correlations were statistically significant at the .05 level. The largest correlation was between the first nine weeks average and the second nine weeks average, $r = .80$, $p < .05$. The weakest correlation was between the second nine weeks average and the Stanford Achievement Test, $r = .21$, $p > .05$.

Insert Table 4 About Here

Table 5 contains the unrotated factor matrix and communality estimates for the analysis of students in grade 3. In this analysis only one factor was extracted using the eigenvalue-one criterion. The loadings on the factor range from a maximum of .86 for the first nine weeks

average to a low of .59 for the Stanford Achievement Test. The communality estimates ranged from a high of .73 for the first nine weeks average to a low of .35 for the Stanford Achievement Test. The table also shows that the single factor accounted for 59.6% of the variability. Because only one factor was extracted, the Varimax rotation was not applied.

Insert Table 5 About Here

Analyses for Students in Grades 4 through 6

The correlations between the mathematics measures for students in grades 4 through 6 are contained in Table 6. The table illustrates that eleven of the fifteen pairwise correlations were statistically significant at the .05 level. The strongest correlation was between the first and second nine weeks averages, $r = .64$, $p < .05$. The weakest correlation was between the first nine weeks average and the Stanford Achievement Test, $r = -.02$, $p < .05$.

Insert Table 6 About Here

Table 7 contains the unrotated factor solution and communalities for the analysis involving students in grades 4 through 6. The table shows that two factors were extracted using the same criterion as the previous two analyses. The unrotated loadings for factor 1 range from a high of .84 for the first nine weeks average to a low of .11 for the Stanford Achievement Test. For factor 2, the loadings range from a high of .92 for the third nine weeks average to .18 for the first nine weeks average. The table also shows that the communalities range from a high of .86 for the third nine weeks average to a low of .38 for the Stanford Achievement Test and that

factor 1 accounts for 38.3% of the variability while factor 2 accounts for an additional 31.8% of the variability.

Insert Table 7 About Here

The results of the Varimax rotation technique are contained in Table 8. The table shows that four of the six variables show a significant association with only one of the two factors. The two remaining variables, first and second nine weeks averages, do show significant associations with factor 1, .78 and .71, respectively; however, they also show associations with factor 2, .36 and .58, respectively.

Insert Table 8 About Here

The high rotated factor loading for the CTB Mathematics Portfolio and the low factor loading for the Stanford Achievement Test on factor 1 suggests that this factor represents constructs that are measured by the CTB Mathematics Portfolio but not the Stanford Achievement Test. Likewise, the factor loadings of the two variables on factor 2 suggest that this factor represents constructs that are associated with the Stanford Achievement Test opposed to the CTB Mathematics Portfolio. An examination of factor 1 in terms of the quarterly classroom assessments shows significant associations for quarters 1, 2, and 4 with only a slight decrease in size from quarters 1 to 2. Despite the strong associations of the first two and last quarters, the association between quarter 3 and factor 1 is nearly zero.

The factor loadings for the classroom assessment on factor 2 also show a noticeable pattern. There appears to be an increasing trend from quarter 1 to quarter 3. However, the trend ends at quarter 3, and the magnitude of the loading drops drastically during quarter 4.

In summary, the only analysis that did not yield two factors was the grade 3 sample indicating that for grade 3 the three forms of assessment are aligned and attempt to measure the same construct. However, the results of the remaining two analyses showed that the three forms of assessment represented two underlying constructs. One construct appeared to be related to assessment defined by the Stanford Achievement Test while the other construct was related to assessment in terms of the CTB Mathematics Portfolio. Additionally, the percent of variance accounted for by each of the factors indicates that neither factor served as a predominant underlying construct.

The results also indicate that classroom assessment in general was related to both factors supporting a statement by the Mathematical Sciences Education Board (1993) who said that teachers may attempt to prepare students for tests while simultaneously offering a richer and more challenging curriculum. Although classroom assessment was related to both factors, distinct patterns among the classroom assessments can be identified. In both analyses, the association between classroom assessment and the factor representing the Stanford Achievement Test increased until quarter three, when the Stanford Achievement Test was administered.

Conclusions

There are several limitations involved in this study that must be considered when viewing the conclusions. First, the sample involved in this study was nonrandom which affects the

generalizability of the results. Therefore, they should be viewed as a description of the consistency and relationships that existed for this sample of students. The second limitation concerns the reliability and validity of the dependent variables, particularly the CTB Mathematics Portfolio since the scoring was modified. Third, the sample sizes used for the grade analyses were below the five subjects per variable minimum suggested by Hair et al. However, the overall sample was adequate using the same criteria. Finally, because the CTB Mathematics Portfolio was only administered during the fourth nine weeks, it would be difficult to determine the patterns that would have resulted if the assessment was used throughout the school year as it was designed.

Based on the results of the three analyses, three conclusions were reached. The first conclusion supports statements by Kulm (1990), Cain and Kennedy (1992), and Romberg and Wilson (1992) who said that classroom assessment is influenced by the tests that the students are mandated to take. The next conclusion is based on the overall and grade 4 through 6 samples which indicate that the Stanford Achievement Test and the CTB Mathematics Portfolio assess different constructs. Finally, based on the grade 3 and grade 4 through 6 analyses, it appears that the constructs assessed by the Stanford Achievement Test change as students move from grade 3 to grade 4.

References

Cain, R. W., & Kennedy, P. A. (1992, November). A joint vision for classroom assessment. The Mathematics Teacher, 85(8), 612-615.

CTB portfolio assessment: Integrated mathematics: Administration manual: Grade 3. (1992). Monterey, CA: CTB Macmillan/McGraw-Hill.

Hair, J. F., Jr., Anderson, R. E., Tatham R. L., & Black, W. C. (1995). Multivariate data analysis with readings (4th ed.). Englewood Cliffs, CA: Prentice Hall.

Kulm, G. (Ed.). (1990). Assessing higher order thinking in mathematics. Washington, DC: American Association for the Advancement of Science.

Mathematical Sciences Education Board. (1993). Measuring what counts: A conceptual guide for mathematics assessment. Washington, DC: National Academy Press.

Romberg, T. A., & Wilson, L. D. (1992, September). Alignment of tests with the standards. Arithmetic Teacher, 40(1), 18-22.

Stanford achievement test series, eighth edition: Technical data report. (1990). Harcourt Brace Jovanovich, Inc.

BEST COPY AVAILABLE

Table 1

Correlations Among the Mathematics Measures for the Overall Sample

	SAT	Portfolio	1st 9 weeks	2nd 9 weeks	3rd 9 weeks	4th 9 weeks
SAT	1.0000					
Portfolio	.12232 (.19871)	1.0000				
1st 9 weeks	.18829 (.09519)	.36831 (.00425)	1.0000			
2nd 9 weeks	.23765 (.04827)	.38819 (.00267)	.70698 (.000)	1.0000		
3rd 9 weeks	.38961 (.00258)	.03314 (.40963)	.33196 (.00926)	.43456 (.00081)	1.0000	
4th 9 weeks	.14878 (.15124)	.41850 (.00125)	.18168 (.10334)	.28383 (.02289)	-.12999 (.18413)	1.0000

Table 2

Unrotated Factor Solution, Communalities and Percent Variance for the Overall Sample

	Factor 1	Factor 2	communality
Stanford Achievement Test	.47921	-.35451	.35533
CTB Mathematics Portfolio	.60551	.52706	.64444
First 9 Weeks Average	.80342	-.03955	.64705
Second 9 Weeks Average	.86567	-.04803	.75169
Third 9 Weeks Average	.53804	-.69436	.77162
Fourth 9 Weeks Average	.43880	.67843	.65280
Percent of Variance	41.2	22.5	

Table 3

Varimax Rotated Factor Loadings for the Overall Sample

	Factor 1	Factor 2
Stanford Achievement Test	.59550	.02650
CTB Mathematics Portfolio	.13832	.79076
First 9 Weeks Average	.64892	.47534
Second 9 Weeks Average	.70262	.50796
Third 9 Weeks Average	.85525	-.20041
Fourth 9 Weeks Average	-.08651	.80332

Table 4

Correlations Among the Mathematics Measures for Grade 3 Students

	SAT	Portfolio	1st 9 weeks	2nd 9 weeks	3rd 9 weeks	4th 9 weeks
SAT	1.0000					
Portfolio	.33688 (.06263)	1.0000				
1st 9 weeks	.44925 (.01798)	.55158 (.00389)	1.0000			
2nd 9 weeks	.21486 (.16846)	.54997 (.004)	.79958 (.000)	1.0000		
3rd 9 weeks	.44536 (.01890)	.45464 (.01676)	.55408 (.00373)	.50418 (.00836)	1.0000	
4th 9 weeks	.46328 (.01495)	.49342 (.00981)	.54218 (.00457)	.66850 (.00034)	.59417 (.00177)	1.0000

Table 5

Unrotated Factor Loadings for Grade 3 Students

	Factor 1	communality
Stanford Achievement Test	.59426	.35315
CTB Mathematics Portfolio	.73112	.53453
First 9 Weeks Average	.85614	.73297
Second 9 Weeks Average	.83387	.69535
Third 9 Weeks Average	.76773	.58941
Fourth 9 Weeks Average	.82001	.67242
Percent of Variance	59.6	

Table 6

Correlations Among the Mathematics Measures for Grade 4 through 6 Students

	SAT	Portfolio	1st 9 weeks	2nd 9 weeks	3rd 9 weeks	4th 9 weeks
SAT	1.0000					
Portfolio	-.21516 (.13577)	1.0000				
1st 9 weeks	-.01926 (.46125)	.36896 (.02667)	1.0000			
2nd 9 weeks	.28493 (.07083)	.31952 (.04871)	.64495 (.00011)	1.0000		
3rd 9 weeks	.36612 (.02767)	-.36388 (.02849)	.32484 (.04584)	.43669 (.01008)	1.0000	
4th 9 weeks	.02980 (.44017)	.48252 (.00465)	.42166 (.01271)	.33218 (.04208)	-.41034 (.01505)	1.0000

Table 7

Unrotated Factor Solution, Communalities and Percent Variance for Grade 4 through 6 Students

	Factor 1	Factor 2	communality
Stanford Achievement Test	.10916	.60769	.38121
CTB Mathematics Portfolio	.64621	-.53731	.70629
First 9 Weeks Average	.84246	.17801	.74143
Second 9 Weeks Average	.81817	.40644	83459
Third 9 Weeks Average	.11802	.92239	.86472
Fourth 9 Weeks Average	.68892	-.45068	.67772
Percent of Variance	38.3	31.8	

Table 8

Varimax Rotated Factor Loadings for Grade 4 Through 6 Students

	Factor 1	Factor 2
Stanford Achievement Test	-.02902	.61674
CTB Mathematics Portfolio	.74970	-.37978
First 9 Weeks Average	.78161	.36129
Second 9 Weeks Average	.70701	.57856
Third 9 Weeks Average	-.09052	.92549
Fourth 9 Weeks Average	.77203	-.28581